

Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format

Akira R. Kinjo¹, Hiroyuki Suzuki¹, Reiko Yamashita¹, Yasuyo Ikegawa¹,
Takahiro Kudou¹, Reiko Igarashi¹, Yumiko Kengaku¹, Hasumi Cho¹,
Daron M. Standley², Atsushi Nakagawa¹ and Haruki Nakamura^{1,*}

¹Institute for Protein Research and ²Immunology Frontier Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received August 8, 2011; Accepted September 15, 2011

ABSTRACT

The Protein Data Bank Japan (PDBj, <http://pdbj.org>) is a member of the worldwide Protein Data Bank (wwPDB) and accepts and processes the deposited data of experimentally determined macromolecular structures. While maintaining the archive in collaboration with other wwPDB partners, PDBj also provides a wide range of services and tools for analyzing structures and functions of proteins, which are summarized in this article. To enhance the interoperability of the PDB data, we have recently developed PDB/RDF, PDB data in the Resource Description Framework (RDF) format, along with its ontology in the Web Ontology Language (OWL) based on the PDB mmCIF Exchange Dictionary. Being in the standard format for the Semantic Web, the PDB/RDF data provide a means to integrate the PDB with other biological information resources.

INTRODUCTION

As a member of the worldwide Protein Data Bank (wwPDB, <http://www.wwpdb.org/>) (1), the Protein Data Bank Japan (PDBj, <http://pdbj.org/>) (2) accepts and processes PDB entries deposited mainly from Asia and Oceania regions and maintains a centralized archive of macromolecular structures, in collaboration with other wwPDB members, RCSB-PDB (3), BMRB (4) in the USA and PDBe (5) in Europe. The PDBj provides its own data viewer, integrated tools and derived databases in order to facilitate structural biology and bioinformatics

research from different perspectives than other members of the wwPDB.

PDBj has been engaged in validation of the PDB data description by developing a canonical XML description (PDBML), primarily in collaboration with RCSB-PDB (6). Based on PDBML, PDBj has established its own data interface, PDBj Mine (<http://pdbj.org/mine>), which enables the user to perform complex queries using SQL, in addition to standard queries (7). The entire relational database is now available for download with weekly updates. RESTful web services are also available for PDBj Mine.

Very recently, in addition to the above, PDBj has begun to provide data in the Resource Description Framework (RDF) format in which most entities are dereferenceable by unique Uniform Resource Identifiers (URI); (See <http://pdbj.org/rdf>). Providing a unique URI to each category element of PDB entries in the RDF format implies that the entire PDB data are accessible on the Web, and that the category elements are linked to one another as well as to external resources. A complete ontology of the RDF in Web Ontology Language (OWL) is also provided, which is based on the established PDB mmCIF Exchange Dictionary so that specialists in structural biology or structural bioinformatics can readily understand the data structures. We believe that this service provides a useful means for automatically processing complex structural information. More importantly, this system should make it possible for the structural information in the PDB to be readily integrated with other biological information resources using Semantic Web frameworks. This will facilitate integration with the new Japanese National Bioscience Database Center (NBDC, <http://biosciencedbc.jp/?l=eng>) founded in April 2011.

*To whom correspondence should be addressed. Tel: +81 6 6879 4311; Fax: +81 6 6879 8636; Email: harukin@protein.osaka-u.ac.jp

WEB INTERFACE

PDBj Mine for searching and retrieving PDB data

The top page of the PDBj website (<http://pdbj.org>) is provided in English, Japanese, Korean and Chinese (simplified and traditional). Tutorials for data deposition are also provided in these languages.

The main query interface is PDBj Mine (7), which is based on a relational database (RDB) derived from PDBMLplus files. PDBMLplus files are an enhancement of PDBML files (6) with additional annotations that have been manually curated or automatically extracted from PDB atomic coordinates or other data sources such as UniProt (8). The search box on the top page can handle general keywords, author names and amino acid sequence fragments as input. A simple search returns a list of PDB entries. Each PDB entry has its summary page from which more detailed pages containing structural, functional and experimental information are linked.

More detailed conditional searches can be performed using the PDBj Mine advanced search interface with which the user can specify a pre-defined set of categories such as release/deposition date, citations, chain types, compounds, experimental techniques, etc.

PDBj Mine also allows searching the entire database by using the SQL interface. If the user is familiar with the terms of the PDB mmCIF Exchange Dictionary (<http://mmcif.pdb.org/dictionaries/>), most searches can be accomplished by using category views. The details of the database structure and tutorial for SQL queries can be found in our previous publication (7) as well as in the PDBj Help page (<http://pdbj.org/doc/help>).

The entire RDB is made available for download at our FTP site (<ftp://ftp.pdbj.org/mine>). The database dump file is weekly updated and incremental update files are also available.

RESTful web services

Two simple RESTful web service interfaces are provided to allow applications to access the search and retrieval services of PDBj Mine through HTTP GET and POST protocols.

One service retrieves a specific element of PDBMLplus data for a PDB entry by specifying an XPath expression in the URL (<http://service.pdbj.org/mine/xpath/>). For example, the HTTP GET protocol to the URL [http://service.pdbj.org/mine/xpath/1GOF//PDBx:entity\[@id='1'\]](http://service.pdbj.org/mine/xpath/1GOF//PDBx:entity[@id='1']) returns the content of the PDBML for the entity category with the item entity.id = '1' of the PDB entry 1GOF.

The other service queries the RDB of PDBj Mine using SQL expressions (<http://service.pdbj.org/mine/sql>). Simply posting an SQL query to this URL returns the results of the query in XML, CSV (comma-separated value) or TSV (tab-separated value) formats. This interface may be a useful means for simple statistical analyses of PDB data.

PDB/RDF: PDB DATA ON THE SEMANTIC WEB

While various web-based services facilitate large-scale use of PDB data (9), their specialized application program interfaces and formats make it difficult to integrate their outputs with those from other resources (10). The Semantic Web is one possible way to overcome this obstacle (11,12) using ‘smart data’ rather than smart applications (13). To make PDB data compatible with the Semantic Web, we have converted PDBML data into the RDF format (14), and made it available on our website (<http://pdbj.org/rdf>). The RDF is a standard format for the Semantic Web (<http://www.w3.org/2011/sw/>) and facilitates exchanging and integrating data originating from diverse sources. Along with the RDF-formatted PDB data (PDB/RDF), we have also translated the PDB mmCIF Exchange Dictionary into the (OWL) (15), which provides the formal semantics of the PDB data elements and enables inference-based queries.

It is noted that the Bio2RDF project (16) has already been providing a subset of PDB data in the RDF format based on the Bio2RDF ontology (<http://bio2rdf.org/>). While Bio2RDF aims at integrating essential subsets of a wide range of biological information resources including the PDB, PDB/RDF aims to complement the Bio2RDF by providing more detailed and more comprehensive information specific to the PDB. In the following, we describe the details of the PDB/RDF and its OWL ontology.

OWL ontology based on the PDB exchange data dictionary

The PDB OWL ontology (<http://pdbj.org/schema/pdbx-v40.owl>) is semi-automatically generated from the PDBML Schema (version 4.0) (6) which in turn is based on the PDB mmCIF Exchange Dictionary (<http://mmcif.pdb.org/dictionaries/>) (17). The PDBML Schema defines not only categories and category items, but also the relationship between categories via primary keys and foreign keys. Thus, the categories in each PDB entry can be linked to each other. In addition, some category items refer to external resources such as sequence databases, biological taxonomy and citations. The PDB OWL ontology accounts for this linked structure of the PDBML Schema.

The basic conversion scheme from the PDBML Schema to the OWL ontology is as follows:

- (1) Each PDBML category corresponds to an OWL class;
- (2) The PDBML ‘datablock’ element representing each PDB entry corresponds to the OWL class ‘datablock’;
- (3) Each datablock instance is related to category element holders (‘xyzCategory’ XML elements in PDBML where ‘xyz’ is a category such as entity, struct, chem_comp, etc.) by OWL properties (‘has_xyzCategory’);
- (4) Each category element holder is related to category elements (‘xyz’ elements in PDBML) by OWL properties (‘has_xyz’);

A

Subject: <http://pdbj.org/rdf/1GOF>

Predicate	Object
PDBo: datablockName	1GOF-noatom
PDBo: has_atom_sitesCategory	PDBr:1GOF/atom_sitesCategory
PDBo: has_atom_sites_footnoteCategory	PDBr:1GOF/atom_sites_footnoteCategory
PDBo: has_atom_typeCategory	PDBr:1GOF/atom_typeCategory
PDBo: has_audit_authorCategory	PDBr:1GOF/audit_authorCategory
PDBo: has_audit_conformCategory	PDBr:1GOF/audit_conformCategory
PDBo: has_cellCategory	PDBr:1GOF/cellCategory
PDBo: has_chem_compCategory	PDBr:1GOF/chem_compCategory
PDBo: has_citationCategory	PDBr:1GOF/citationCategory
PDBo: has_citation_authorCategory	PDBr:1GOF/citation_authorCategory
PDBo: has_computingCategory	PDBr:1GOF/computingCategory
PDBo: has_database_2Category	PDBr:1GOF/database_2Category
PDBo: has_database_PDB_matrixCategory	PDBr:1GOF/database_PDB_matrixCategory
PDBo: has_database_PDB_revCategory	PDBr:1GOF/database_PDB_revCategory
PDBo: has_database_PDB_rev_recordCategory	PDBr:1GOF/database_PDB_rev_recordCategory
PDBo: has_diffmCategory	PDBr:1GOF/diffrnCategory
PDBo: has_diffm_radiationCategory	PDBr:1GOF/diffrn_radiationCategory
PDBo: has_diffm_radiation_wavelengthCategory	PDBr:1GOF/diffrn_radiation_wavelengthCategory
PDBo: has_entityCategory	PDBr:1GOF/entityCategory
PDBo: has_entity_keywordsCategory	PDBr:1GOF/entity_keywordsCategory
PDBo: has_entity_polyCategory	PDBr:1GOF/entity_polyCategory
PDBo: has_entity_src_genCategory	PDBr:1GOF/entity_src_genCategory
PDBo: has_entryCategory	PDBr:1GOF/entryCategory
PDBo: has_exptlCategory	PDBr:1GOF/exptlCategory
PDBo: has_exptl_crystalCategory	PDBr:1GOF/exptl_crystalCategory
PDBo: has_pdbx_database_remarkCategory	PDBr:1GOF/pdbx_database_remarkCategory
PDBo: has_pdbx_database_statusCategory	PDBr:1GOF/pdbx_database_statusCategory
PDBo: has_pdbx_entity_nonpolyCategory	PDBr:1GOF/pdbx_entity_nonpolyCategory
PDBo: has_pdbx_entry_detailsCategory	PDBr:1GOF/pdbx_entry_detailsCategory
PDBo: has_pdbx_struct_assemblyCategory	PDBr:1GOF/pdbx_struct_assemblyCategory
PDBo: has_pdbx_struct_assembly_genCategory	PDBr:1GOF/pdbx_struct_assembly_genCategory
PDBo: has_struct_assembly_refCategory	PDBr:1GOF/struct_assembly_refCategory

B

Subject: <http://pdbj.org/rdf/1GOF/entityCategory>

Predicate	Object
PDBo: of_datablock	PDBr:1GOF
PDBo: has_entity	PDBr:1GOF/entity/1
PDBo: has_entity	PDBr:1GOF/entity/2
PDBo: has_entity	PDBr:1GOF/entity/3
PDBo: has_entity	PDBr:1GOF/entity/4
PDBo: has_entity	PDBr:1GOF/entity/5
rdf:type	http://pdbj.org/schema/pdbx-v40.owl#entityCategory

C

Subject: <http://pdbj.org/rdf/1GOF/entity/1>

Predicate	Object
PDBo: of_datablock	PDBr:1GOF
PDBo: of_category	PDBr:1GOF/entityCategory
PDBo: entity.formula_weight	68579.250
PDBo: entity.id	1
PDBo: entity.pdbx_description	GALACTOSE OXIDASE
PDBo: entity.pdbx_ec	1.1.3.9
PDBo: entity.pdbx_number_of_molecules	1
PDBo: entity.src_method	man
PDBo: entity.type	polymer
PDBo: link_to_enzyme	http://purl.uniprot.org/enzyme/1.1.3.9
PDBo: referenced_by_entity_keywords	PDBr:1GOF/entity_keywords/1
PDBo: referenced_by_entity_poly	PDBr:1GOF/entity_poly/1
PDBo: referenced_by_entity_src_gen	PDBr:1GOF/entity_src_gen/1
PDBo: referenced_by_struct_asym	PDBr:1GOF/struct_asym/A
PDBo: referenced_by_struct_ref	PDBr:1GOF/struct_ref/1
rdf:type	http://pdbj.org/schema/pdbx-v40.owl#entity

Figure 1. Example of PDB/RDF pages. (A) The PDB/RDF page for the PDB entry 1GOF (<http://pdbj.org/rdf/1GOF>) displays a list of RDF triples whose subject is the URL itself and objects are mostly the links to the category element holders. (B) The PDB/RDF page for the ‘entity’ category element holder (<http://pdbj.org/rdf/1GOF/entityCategory>) of the same entry displays a list of entity category elements. (C) The PDB/RDF page for the entity category with entity.id = 1 (<http://pdbj.org/rdf/1GOF/entity/1>) displays the properties (category items) of the element as well as to links to external resources and cross-references within the entry. The source files of these pages are in the RDF/XML format.

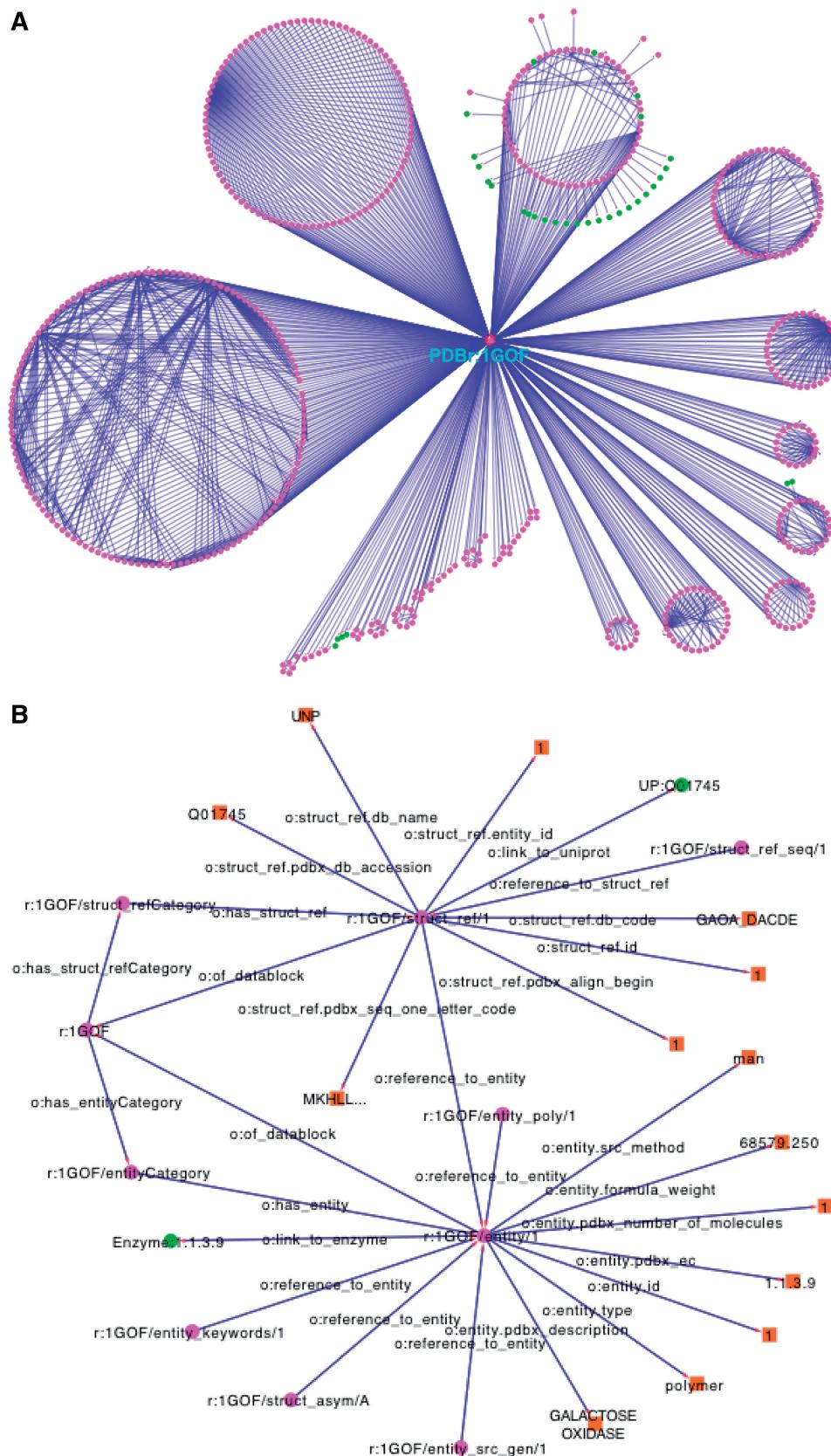


Figure 2. Example of an RDF graph for PDB entry 1GOF. (A) The network of RDF resources for the PDB entry 1GOF. Nodes in magenta represent dereferenceable RDF resources within the PDB entry, nodes in green represent external resources; edges represent links between nodes. (B) A subgraph of (A) augmented with literal objects (orange rectangles). Nodes are labeled with simplified URL or literal values and edges with predicates defined in the OWL ontology ('r:' and 'o:' are short-hands for <http://pdblj.org/rdf/> and [http://pdblj.org/schema/pdbx-v40.owl#/](http://pdblj.org/schema/pdbx-v40.owl#), respectively; 'UP:' and 'Enzyme:' indicate <http://purl.uniprot.org/uniprot/> and <http://purl.uniprot.org/enzyme/>, respectively). Diagrams were created using Cytoscape (21).

Table 1. Analysis tools and derived databases

Service	Description	URL ^a	Reference
PDBj Mine	Simple, advanced and SQL search for PDB data	pdbj.org/mine	(7)
Sequence Navigator	BLAST (22) search against PDB	pdbj.org/seqnavix	—
Structure Navigator	ASH structure similarity search	pdbj.org/stnavix	(23)
SeSAW	Functional site prediction	pdbj.org/sesaw	(24)
GIRAF	Interaction interface similarity search	pdbj.org/giraf	(25–27)
EM Navigator	EM structure browser	pdbj.org/emnavi	—
jV	Molecular graphics viewer	pdbj.org/jv	(28)
Yorodumi	Integrated PDB/EMDB data browser	pdbj.org/yorodumi	—
Protein Globe	“Protein universe” browser	pdbj.org/globe	(2)
ASH	Pairwise structure alignment	pdbj.org/ash	(23,29,30)
MAFTTash	Integrated multiple sequence/structure alignment	pdbj.org/mafftash	(31)
CRNPRED	One-dimensional structure prediction	pdbj.org/crnpred	(32,33)
Spanner	Homology modeling	pdbj.org/spanner	(34)
SFAS	Automated function annotation	pdbj.org/sfas	—
eF-site	Electrostatic surface database	pdbj.org/ef-site	(28,35)
eF-seek	Search against eF-site	pdbj.org/ef-seek	(36)
eF-surf	Building electrostatic surfaces	pdbj.org/ef-surf	(36)
ProMode	Database of normal mode analyses of proteins	pdbj.org/promode	(37)
ProMode Elastic	ProMode with elastic network model	pdbj.org/promode_elastic	—
ProMode Oligomer	ProMode for oligomeric proteins	pdbj.org/promode_oligomer	—
eProtS	Educational articles on protein structures	pdbj.org/eports	—
Molecule of the Month	Japanese translation of RCSB PDB MoM	pdbj.org/mom	(3)

^aThe ‘http://’ prefix is omitted.

- (5) When two categories are related by a foreign key in the PDBML Schema [by means of key and keyref elements of the XML Schema (18)], a corresponding OWL property is defined for cross-references (e.g. ‘reference_to_xyz’ and ‘referenced_by_xyz’);
- (6) When a category item refers to an external resource, that item is linked to the resource by a property (e.g. ‘link_to_uniprot’).
- (7) Category items are defined as datatype properties. For example, the ‘entity.pdbx_ec’ item of the entity category is translated to the ‘entity.pdbx_ec’ property.

Thus, the PDB/RDF ontology is a straightforward translation of the PDB mmCIF Exchange Dictionary from which users familiar with the dictionary can easily infer the meaning of and interrelationship between the PDB/RDF data. In addition, the PDB/RDF ontology defines simple hierarchical structures for classes and properties so that, in principle, related categories or category items can be obtained at once based on the OWL inference mechanism.

RDF data based on PDBML

The RDF files of PDB data are automatically generated from PDBML files by XSL transformation (19). However, to reduce the data size, some categories, namely atom_site, entity_poly_seq, pdbx_poly_seq_scheme and pdbx_nonpoly_scheme, are currently excluded. The user who need the full set of RDF data may easily generate them by applying the XSLT stylesheet provided at <http://pdbj.org/rdf> to PDBML files. As of August 3, there are in total 565 236 618 RDF triples for 74 888 PDB entries stored in the PDB/RDF.

Each category element is assigned a URL that is actually accessible on the Web. For example (Figure 1), the datablock of the PDB entry 1GOF (20) can be accessed at <http://pdbj.org/rdf/1GOF>, which contains mainly a list of RDF triples whose subject is the given URL (i.e. <http://pdbj.org/rdf/1GOF>) and whose objects are mostly the URLs of the category holders. Its category holder for the entity category is found at <http://pdbj.org/rdf/1GOF/entityCategory>, which in turn contains a list of RDF triples whose subject is the given URL (i.e. <http://pdbj.org/rdf/1GOF/entityCategory>) and whose objects are mainly its category elements. In general, the URL of a category holder ‘xyzCategory’ for a PDB entry is provided as <http://pdbj.org/rdf/PDBID/xyzCategory>. A particular entity category element with entity.id = 1 is found at <http://pdbj.org/rdf/1GOF/entity/1>, which lists the content of category items as well as links, if available, to external resources and other category elements in the same entry. Note that the item entity.id is the primary key for the entity category. In general, the URL of a category element ‘xyz’ is provided as <http://pdbj.org/rdf/PDBID/xyz/pkey1,pkey2,...> where pkey1, pkey2, ... are the values of the category items that comprise the primary key. In each of these pages, an RDF/XML format data of the subject, property, object triples with its subject being the one specified by the URL is returned. As a whole, the PDB/RDF data of each PDB entry comprise a network of Web resources that are accessible with explicit URLs and are also connected to external Web resources (Figure 2).

Web interface to PDB/RDF

Although the PDB/RDF data are primarily intended for software agents, they are also useful for human users to

A

The Yorodumi interface displays a complex 3D molecular structure composed of multiple colored subunits (blue, green, red, orange) against a grey background. The top menu bar includes 'Options' and '日本語'. A dropdown menu shows 'PDB-3mfp' and 'Jmol'. The main window title is 'Jmol'. To the right are three floating windows: 'Data' (Database ID: PDB-3mfp, Title: 'Atomic model of F-actin based on a 6.6 angstrom resolution cryoEM map Actin, alpha skeletal muscle'), 'View' (Viewer settings: Size, Higher quality (slower), Zoom Small to Large), and 'Help'.

B

The EM Navigator interface features a '3D Electron Microscopy (3D-EM) Data Navigator [English / 日本語]' header. The top navigation bar includes 'Top', 'Gallery', 'List', 'Diagram', 'Viewer', and 'Documents'. Below is a search bar and a filter section for 'Order: Category', 'Image: Rainbow', and 'Size: Small'. The main area is a 'Gallery page' displaying a grid of 3D electron microscopy density maps. The first row is titled 'cell cycle ...' and the second row is titled 'DNA replication & repair'. Each map is color-coded (rainbow gradient) and includes a scale bar in the bottom left corner.

Figure 3. Yorodumi and EM Navigator. (A) Yorodumi displaying an EMDB-PDB hybrid data, EMDB-5168 and PDB 3MFP (helical assembly) (40). (B) EM Navigator gallery page.

explore the relationship between different data elements in PDB entries. Thus, we provide a simple interface for browsing the PDB/RDF data by applying an XSL stylesheet to the machine-readable RDF format. The user can query the PDB/RDF database with PDB ID, OWL properties or keywords, and follow the links to PDB/RDF resources as well as to external HTML or RDF resources (Figure 1).

OTHER TOOLS AND RESOURCES

In addition to the services described above, PDBj provides a number of analysis tools and derived databases (Table 1). These include search services for various aspects of structural similarity (Sequence Navigator, Structure Navigator, GIRAF, eF-seek), function prediction and annotation (SeSAW, SFAS), structure prediction and modeling services (CRNPRED, Spanner), sequence and structure alignment (ASH, MAFFTash), derived databases (EM Navigator, eF-site, ProMode) and educational resources (Protein Globe, eProtS, Japanese translation of the Molecule of the Month). Among them, we describe below two recently developed services, Yorodumi and EM Navigator.

Yorodumi and EM Navigator: integration with electron microscopy structures

In addition to the simple summary page for each entry, PDBj also provides a more feature-rich entry-wise interface called Yorodumi (Figure 3A, <http://pdbj.org/yorodumi/>). Yorodumi is an interactive and integrated interface for browsing 3D structure data not only in the PDB but also in the EMDB (Electron Microscopy Data Bank, <http://www.ebi.ac.uk/pdbe/emdb/>) (38). The user can select either Jmol (39) or jV (28) Java applets for interactive molecular graphics. In the pages for PDB data entries, the user can easily find biologically relevant parts in the entry as recorded in the PDB data, such as subunits, ligand-binding residues, regions with some specific sequences. Yorodumi also supports displaying ‘biological assembly’ including ‘split entries’ (large structures which are determined in a single experiment, but are released as multiple PDB entries due to the limitations of the legacy PDB format). In the EMDB entry pages, the user can interactively view polygon representations of the 3D map iso-surfaces. For the EMDB-PDB hybrid data, which carries atomic models fitted to the EM map data, Yorodumi can show the superimposed view of the data from the both databanks.

Since electron microscopy is increasingly becoming important as a structure determination method, especially for very large molecular complexes, we also provide EM Navigator, a tool that is specifically designed to explore electron microscopy (EM) structure data deposited in the EMDB (38). To view EMDB 3D map data easily, movies of 3D structure models rolling along horizontal and vertical axes are provided. They are shown in iso-surface representations for typical 3D map data or in solid representations for noisy structure data such as electron tomograms. When atomic models are available in the PDB,

they are superimposed on lower resolution EM maps. In the details page of EMDB entries, one or more movies are displayed together with metadata information. In the results of keyword searches, the user can customize the kinds of displayed information such as names of the microscopy or software packages used in the experiments in tabular format. In the gallery page, the user can see the snapshot images of all the released EM structures at a glance (Figure 3B).

ACKNOWLEDGEMENTS

We thank Vladimir Guranovic and John Westbrook for helping with the PDBML schema.

FUNDING

Funding for open access charge: National Bioscience Database Center, the Japan Science and Technology Agency.

Conflict of interest statement. None declared.

REFERENCES

- Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
- Standley,D.M., Kinjo,A.R., Kinoshita,K. and Nakamura,H. (2008) Protein structure databases with new web services for structural biology and biomedical research. *Brief. Bioinformatics*, **9**, 276–285.
- Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlić,A., Quesada,M., Quinn,G.B., Westbrook,J.D. et al. (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
- Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. et al. (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
- Velankar,S., Alhroub,Y., Ile Alili,A., Best,C., Boutselakis,H.C., Caboche,S., Conroy,M.J., Dana,J.M., van Ginkel,G., Golovin,A. et al. (2011) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **39**, D402–D410.
- Westbrook,J., Ito,N., Nakamura,H., Henrick,K. and Berman,H.M. (2005) PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*, **21**, 988–992.
- Kinjo,A.R., Yamashita,R. and Nakamura,H. (2010) PDBj Mine: design and implementation of relational database interface for Protein Data Bank Japan. *Database*, **2010**, baq021, doi: 1093/database/baq021.
- Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R., Magrane,M. et al. (2005) The universal protein resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
- Katayama,T., Arakawa,K., Nakao,M., Ono,K., Aoki-Kinoshita,K.F., Yamamoto,Y., Yamaguchi,A., Kawashima,S., Chun,H.W., Aerts,J. et al. (2010) The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows. *J. Biomed. Semantics*, **1**, 8.
- Katayama,T., Wilkinson,M.D., Vos,R., Kawashima,T., Kawashima,S., Nakao,M., Yamamoto,Y., Chun,H.-W., Yamaguchi,A., Kawano,S. et al. (2011) The 2nd DBCLS BioHackathon: interoperable bioinformatics Web services for integrated applications. *J. Biomed. Semantics*, **2**, 4.
- Good,B.M. and Wilkinson,M.D. (2006) The life sciences semantic web is full of creeps! *Brief. Bioinform.*, **7**, 275–286.

12. Pasquier,C. (2008) Biological data integration using Semantic Web technologies. *Biochimie*, **584**, 584–594.
13. Allemand,D. and Handler,J. (2008) *Semantic Web for the Working Ontologist*. Morgan Kaufmann, Burlington, MA, USA.
14. Manola,F. and Miller,E. (eds) (2004) RDF Primer, World Wide Web Consortium. <http://www.w3.org/TR/rdf-primer> (27 September 2011, date last accessed).
15. McGuinness,D.L. and van Harmelen,F. (eds) (2009) OWL Web Ontology Language Overview, World Wide Web Consortium. <http://www.w3.org/TR/owl-features/> (27 September 2011, date last accessed).
16. Belleau,F., Nolin,M.A., Tourigny,N., Rigault,P. and Morissette,J. (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform.*, **41**, 706–716.
17. Westbrook,J.D. and Bourne,P.E. (2000) STAR/mmCIF: an ontology for macromolecular structure. *Bioinformatics*, **159**, 159–168.
18. Fallside,D.C. and Walmsley,P. (eds) (2004) XML Schema Part 0: Primer, World Wide Web Consortium. <http://www.w3.org/TR/xmlschema-0/> (27 September 2011, date last accessed).
19. Clark,J. (ed.) (1999) XSL Transformations (XSLT) Version 1.0, World Wide Web Consortium. <http://www.w3.org/TR/xslt/> (27 September 2011, date last accessed).
20. Ito,N., Phillips,S.E., Stevens,C., Ogel,Z.B., McPherson,M.J., Keen,J.N., Yadav,K.D. and Knowles,P.F. (1991) Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature*, **350**, 87–90.
21. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
22. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.L. (1997) Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
23. Standley,D.M., Toh,H. and Nakamura,H. (2007) ASH structure alignment package: sensitivity and selectivity in domain classification. *BMC Bioinformatics*, **8**, 116.
24. Standley,D.M., Yamashita,R., Kinjo,A.R., Toh,H. and Nakamura,H. (2010) SeSAW: balancing sequence and structural information in protein functional mapping. *Bioinformatics*, **26**, 1258–1259.
25. Kinjo,A.R. and Nakamura,H. (2007) Similarity search for local protein structures at atomic resolution by exploiting a database management system. *Biophysics*, **3**, 75–84.
26. Kinjo,A.R. and Nakamura,H. (2009) Comprehensive structural classification of ligand binding motifs in proteins. *Structure*, **17**, 234–246.
27. Kinjo,A.R. and Nakamura,H. (2010) Geometric similarities of protein-protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study. *J. Mol. Biol.*, **399**, 526–540.
28. Kinoshita,K. and Nakamura,H. (2004) eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics*, **20**, 1329–1330.
29. Standley,D.M., Toh,H. and Nakamura,H. (2005) GASH: an improved algorithm for maximizing the number of equivalent residues between two protein structures. *BMC Bioinformatics*, **6**, 221.
30. Standley,D.M., Toh,H. and Nakamura,H. (2004) Detecting local structural similarity in proteins by maximizing number of equivalent residues. *Proteins*, **57**, 381–391.
31. Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.*, **9**, 286–298.
32. Kinjo,A.R. and Nishikawa,K. (2005) Predicting secondary structures, contact numbers, and residue-wise contact orders of native protein structure from amino acid sequence using critical random networks. *Biophysics*, **1**, 67–74.
33. Kinjo,A.R. and Nishikawa,K. (2006) CRNPRED: highly accurate prediction of one-dimensional protein structures by large-scale critical random networks. *BMC Bioinformatics*, **7**, 401.
34. Lis,M., Kim,T., Sarmiento,J., Kuroda,D., Dinh,H., Kinjo,A.R., Amada,K., Devadas,S., Nakamura,H. and Standley,D.M. (2011) Bridging the gap between single-template and fragment based protein structure modeling using Spanner. *Immunome Res.*, **7**, 1–8.
35. Kinoshita,K. and Nakamura,H. (2003) Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.*, **12**, 1589–1595.
36. Kinoshita,K., Murakami,Y. and Nakamura,H. (2007) eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape. *Nucleic Acids Res.*, **35**, W398–W402.
37. Wako,H., Kato,M. and Endo,S. (2004) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, **20**, 2035–2043.
38. Lawson,C.L., Baker,M.L., Best,C., Bi,C., Dougherty,M., Feng,P., van Ginkel,G., Devkota,B., Lagerstedt,I., Ludtke,S.J. et al. (2010) EMDDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
39. Hanson,R.M. (2010) Jmol a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.*, **43**, 1250–1260.
40. Fujii,T., Iwane,A.H., Yanagida,T. and Namba,K. (2010) Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, **467**, 724–728.